Floating Ground-truth Labels

Jiman Kim
Samsung Research
jiman14.kim@samsung.com

Chanjong Park
Samsung Research
cj710.park@samsung.com

Abstract

The reliability of ground-truth labels is very important in supervised image classification problems. Although a large amount of images are collected, if the amount of noisy labels is high, deep neural networks cannot learn the discriminative visual properties between the target classes; this results in underfitting. In real-world applications, this situation frequently occurs. Regardless of the number of noise labels because of unclear visual patterns and subjective (floating) labeling criterion, deep networks should maintain high performance. In this paper, we describe this problem in detail and introduce a related application. To apply the existing deep learning based algorithms to large-scale, real-world image data, it is important to address the noisy label problem.

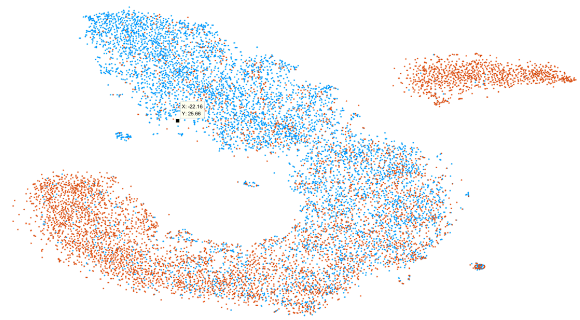

Figure 1: Data distribution with high ratio of noisy labels in feature space. The locations of data points are computed by the t-SNE method using a pretrained convolutional deep network. Blue and red points represent different classes.

1. Introduction

A necessary condition for supervised image classification is to clearly define the target objects and assign accurate class labels for the acquired images. To improve the classification accuracy, we should obtain a large amount of images, including many variations of objects and backgrounds. However, if incorrect labels are assigned to a high-quality image dataset, the ground-truth of each label represents a random distribution in the feature space. This results in the underfitting of deep neural networks.

In a general labeling process, rough labels are assigned through automatic preprocessing algorithms and final labels are manually decided. Thus, the clearer the classification criteria between objects, the more reliable is the ground-truth set. However, unlike object image datasets with unique structural information, non-object image datasets consist of ambiguous non-rigid patterns. In such cases, although a cross-checking is performed by multiple human inspectors, many incorrect labels exist in the ground-truth dataset because the degree of the labeling consistency for the same image is low. This problem is caused by various reasons, and it is difficult to directly address them. In addition, it is difficult to train deep networks using a consistent strategy because the rate at which label noise occupies the data also varies. We explain the detailed reasons that generate severe label noise in both the training and test dataset considering a specific real-world application, such as a high-level visual inspection.

2. Floating Ground-truth Labels

2.1. Ambiguous Visual Patterns

There are three cases that increase the amount of falsely assigned labels in the ground-truth dataset. First, when the visual pattern between different classes is ambiguous. As mentioned in Section 1, this occurs when the components of the target classes are random patterns and not objects. As a result, the labeling criteria are floating, not fixed. For instance, the visual patterns of the real defect images and fake defect images [2] are inconsistent, and it is very difficult to classify them into different classes. We found that the data distributions of the two classes overlap considerably in a feature space, which is obtained from t-stochastic neighbor embedding (t-SNE) [7] (Fig. 1). We can decide if the labels of some pattern are dust, stain, or scratch by rubbing the surface at which the patterns appear. This means that we cannot assign the true labels only by observing images.